

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Selection and environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*.

### Permalink

<https://escholarship.org/uc/item/9w086955>

### Journal

Proceedings of the National Academy of Sciences of the United States of America, 115(22)

### ISSN

0027-8424

### Authors

Wang, Guo-Dong  
Zhang, Bao-Lin  
Zhou, Wei-Wei  
et al.

### Publication Date

2018-05-01

### DOI

10.1073/pnas.1716257115

Peer reviewed

# Selection and environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*

Guo-Dong Wang<sup>a,b,1</sup>, Bao-Lin Zhang<sup>a,c,1</sup>, Wei-Wei Zhou<sup>a,d</sup>, Yong-Xin Li<sup>a,c</sup>, Jie-Qiong Jin<sup>a</sup>, Yong Shao<sup>a</sup>, He-chuan Yang<sup>e</sup>, Yan-Hu Liu<sup>f</sup>, Fang Yan<sup>a</sup>, Hong-Man Chen<sup>a</sup>, Li Jin<sup>g</sup>, Feng Gao<sup>h</sup>, Yaoguang Zhang<sup>g</sup>, Haipeng Li<sup>b,h</sup>, Bingyu Mao<sup>a,b</sup>, Robert W. Murphy<sup>a,i</sup>, David B. Wake<sup>j,2</sup>, Ya-Ping Zhang<sup>a,b,2</sup>, and Jing Che<sup>a,b,d,2</sup>

<sup>a</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; <sup>b</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China; <sup>c</sup>Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, Yunnan, China; <sup>d</sup>Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, 05282 Nay Pyi Taw, Myanmar; <sup>e</sup>Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore; <sup>f</sup>Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, China; <sup>g</sup>Key Laboratory of Freshwater Fish Reproduction and Development of the Ministry of Education and Key Laboratory of Aquatic Science of Chongqing, Southwest University School of Life Sciences, Chongqing 400715, China; <sup>h</sup>CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; <sup>i</sup>Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, ON, Canada M5S 2C6; and <sup>j</sup>Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720-3160

Contributed by David B. Wake, April 13, 2018 (sent for review September 17, 2017; reviewed by Scott V. Edwards and Craig Moritz)

Tibetan frogs, *Nanorana parkeri*, are differentiated genetically but not morphologically along geographical and elevational gradients in a challenging environment, presenting a unique opportunity to investigate processes leading to speciation. Analyses of whole genomes of 63 frogs reveal population structuring and historical demography, characterized by highly restricted gene flow in a narrow geographic zone lying between matrilineal West (W) and East (E). A population found only along a single tributary of the Yalu Zangbu River has the mitogenome only of E, whereas nuclear genes of W comprise 89–95% of the nuclear genome. Selection accounts for 579 broadly scattered, highly divergent regions (HDRs) of the genome, which involve 365 genes. These genes fall into 51 gene ontology (GO) functional classes, 14 of which are likely to be important in driving reproductive isolation. GO enrichment analyses of E reveal many overrepresented functional categories associated with adaptation to high elevations, including blood circulation, response to hypoxia, and UV radiation. Four genes, including *DNAJC8* in the brain, *TNNC1* and *ADORA1* in the heart, and *LAMB3* in the lung, differ in levels of expression between low- and high-elevation populations. High-altitude adaptation plays an important role in maintaining and driving continuing divergence and reproductive isolation. Use of total genomes enabled recognition of selection and adaptation in and between populations, as well as documentation of evolution along a stepped cline toward speciation.

gene flow | hybridization | natural selection | population genomics | speciation

speciation (16). However, gene flow can also facilitate evolution and speciation by transferring adaptive genes or generating novel genes (17). A genomic approach has the potential to identify genes that are important for adaptation and speciation, which often evolve more rapidly than other genes (18). Growing numbers of studies have used genome-scale comparisons in phylogeography and speciation [e.g., birds (19–22), fishes (23, 24), mammals (25, 26), reptiles (27, 28)].

## Significance

Central topics in evolutionary biology include uncovering the processes and genetic bases of speciation and documenting environmental adaptations and processes responsible for them. The challenging environment of the Qinghai-Tibetan Plateau (QTP) facilitates such investigations, and the Tibetan frog, *Nanorana parkeri*, offers a unique opportunity to investigate these processes. A cohort of whole-genome sequences of 63 individuals from across its entire range opens avenues for incorporating population genomics into studies of speciation. Natural selection plays an important role in maintaining and driving the continuing divergence and reproductive isolation of populations of the species. The QTP is a natural laboratory for studying how selection drives adaptation, how environments influence evolutionary history, and how these factors can interact to provide insight into speciation.

Speciation, the fundamental phenomenon underlying biodiversity, continues to be a central focus of research in evolutionary biology. Disentangling pattern and process and gaining an understanding of the underlying genetic mechanisms as speciation proceeds are central issues in species biology (1, 2). Most vertebrate species arise by vicariant isolation. This form of speciation takes time and the right combination of stochastic genetic change and natural selection (3, 4). It can occur swiftly under certain circumstances, for example, when populations experience rapid demographic changes (e.g., bottlenecks, expansions) (5, 6) or when they are exposed to ecological shifts (7–9). Speciation becomes difficult to understand when it involves many non-exclusive mechanisms (10). Fortunately, approaches using whole genomes or transcriptomes are opening new research pathways (11–13).

Speciation accompanied by gene flow (i.e., without complete geographical isolation) is thought to be common in animal evolution (14, 15). The role of gene flow in speciation remains controversial because it is often assumed to be an impediment to

Author contributions: W.-W.Z., D.B.W., Y.-P.Z., and J.C. designed research; G.-D.W., Y.-P.Z., and J.C. managed this project; B.-L.Z., J.-Q.J., Y.S., and J.C. collected the samples and performed physiological measurement; B.-L.Z., J.-Q.J., F.Y., and H.-M.C. performed DNA experiments; Y.-X.L., Y.S., and B.M. performed the gene expression experiment; B.-L.Z., H.-C.Y., and Y.-H.L. performed the G-PhoCS analysis and coalescent simulations; B.-L.Z., L.J., and Y.Z. performed the histological experiment; F.G. and H.L. calculated the population-scaled recombination rates; G.-D.W. and B.-L.Z. analyzed the data; and G.-D.W., B.-L.Z., W.-W.Z., R.W.M., D.B.W., Y.-P.Z., and J.C. wrote the manuscript.

Reviewers: S.V.E., Harvard University; and C.M., Australian National University.

The authors declare no conflict of interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

The sequence data reported in this paper have been deposited in the genome sequence archive of Beijing Institute of Genomics, Chinese Academy of Sciences, [gsa.big.ac.cn](#) (accession nos. [CRA000919](#) for *N. parkeri*, and [CRA000918](#) for *N. pleskei*).

<sup>1</sup>G.-D.W. and B.-L.Z. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: [chej@mail.kiz.ac.cn](mailto:chej@mail.kiz.ac.cn), [zhangyp@mail.kiz.ac.cn](mailto:zhangyp@mail.kiz.ac.cn), or [wakelab@berkeley.edu](mailto:wakelab@berkeley.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716257115/-DCSupplemental](#).

Published online May 14, 2018.

How organisms adapt and diversify in high-altitude environments has attracted much attention in recent years (29–31). Because of its heterogeneous topography, inhospitable environment, and complex paleoclimate history, the Qinghai-Tibetan Plateau (QTP) is a natural laboratory for studying adaptation and speciation (32–35). Differences in elevation between valleys and mountaintops often exceed 2,000 m, presenting strong ecological gradients in abiotic variables, such as oxygen partial pressure (36), UV radiation (37), precipitation (38), and ambient temperatures (39). The periodicity of climatic oscillations affects diverse phenomena, such as dispersal, fluctuations in population size, and speciation (32, 40).

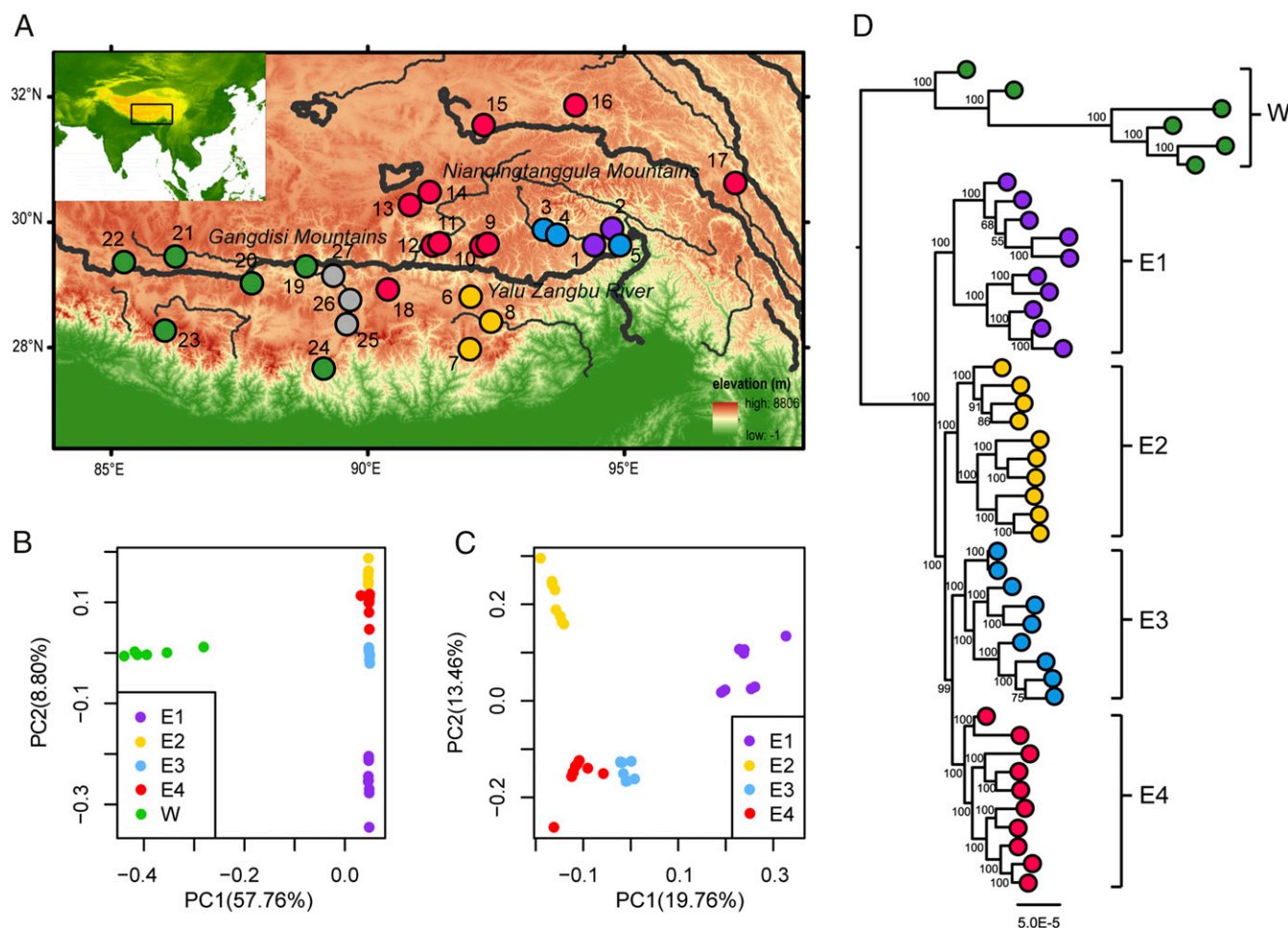
An endemic Tibetan frog, *Nanorana parkeri*, occurs in lentic environments of the southeastern QTP along the Yalu Zangbu River (YZR) drainage. It faces challenges that few other amphibians experience. Not unusual for frogs, it occurs at an elevation of 2,800 m, yet it is the only amphibian to also exist at 5,000 m (41), where oxygen is scarce and UV radiation is dangerously high. Western (W) and eastern (E) mitochondrial DNA matrilineal of *N. parkeri* diverged in the middle Pleistocene and formed distinct entities with overlapping elevational ranges (42). Three nuclear DNA loci correspond to this pattern with limited admixture of E and W alleles in two localities near their geographic boundaries (43). However, neither the extent of gene flow nor the degree of isolation in the zone of

admixture is known. No morphological differences between W and E have been noted (44). Clines in elevation, genetic-morphological discordance, and mtDNA introgression in the zone of admixture offer an unprecedented opportunity to investigate the role selection plays in driving rapid genetic change, environmental adaptation, and the evolution of species differences.

The annotated genome of *N. parkeri* (45) facilitates investigations into the evolutionary drivers of population divergence, and possibly speciation, by enabling the identification of selected genes and their functions. Herein, we report on whole-genome sequences of 63 *N. parkeri* individuals from across the range of the species and decipher the genetic mechanisms that may be responsible for driving population divergence. We reconstruct the evolutionary history of *N. parkeri* from a genomic perspective, investigate the impacts of isolation and gene flow on the process of speciation, and explore genomic consequences. We also evaluate ecological factors leading to speciation and identify candidate genes underlying the observed differentiation.

## Results

**Sampling and Sequencing.** Collection localities ranged from 2,900 to 4,900 m in elevation and covered the entire documented distribution of *N. parkeri* (Fig. 1A and *SI Appendix*, Table S1).

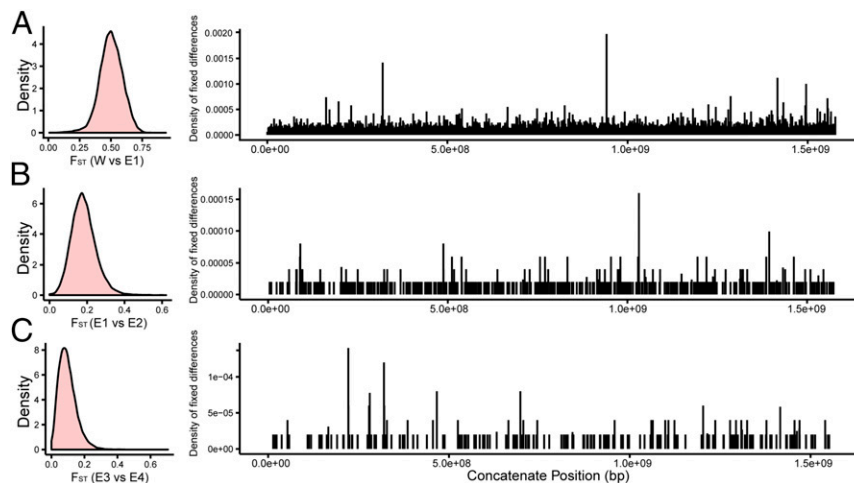


**Fig. 1.** Sampling sites and population structure of the Tibetan frog (*N. parkeri*). (A) Sampling locations (ArcGIS 10.2; esri). Site numbers refer to *SI Appendix*, Table S1. Colors denote the five main groups recovered from population structure and phylogenetic analyses. Gray indicates hybrid populations (25–27). (B) PCA of all high-coverage samples. (C) Principle component plot of E samples only. (D) ML tree based on concatenated sequences. W, green; E1, purple; E2, yellow; E3, light blue; E4, red.









**Fig. 4.** Genomic divergence associated with species formation. (A)  $F_{ST}$  distribution between W and E1 and their landscape of genomic divergence measured by the  $df$ . Both statistics are measured from 50-kb nonoverlapping windows. Scaffolds are concatenated to reveal the whole-genomic divergence pattern. (B)  $F_{ST}$  distribution between E1 and E2 and their landscape of genomic divergence. (C)  $F_{ST}$  distribution between E3 and E4 and their landscape of genomic divergence.

(induction of apoptosis, GO:0006917) were significantly enriched in high-elevation E2 (SI Appendix, Table S10), and response to radiation (GO:0009314) and light stimulus (GO:0009416) were significantly enriched in high-elevation E4 (SI Appendix, Table S11). GO categories associated with blood circulation (GO:0008015) and metabolism process (response to lipid, GO:0033993) are present in low-elevation E3 (SI Appendix, Table S12).

To verify the function of the genes in HDRs, we performed real-time quantitative PCR on seven candidate genes from three different tissues (heart, lung, and brain). These seven candidate genes were identified based upon what they potentially contribute to high-altitude adaptation (SI Appendix, Table S13). Four of these differ in their levels of expression between low-elevation and high-elevation lineages (Fig. 7D), including two related to cardiac function in the heart [Troponin C1 Slow (*TNNC1*) (51) and Adenosine A1 Receptor (*ADORA1*) (52)] and two hypoxia-related genes [DnaJ Homolog Subfamily C Member 8 (*DNAJC8*) in the brain (53) and Laminin Subunit Beta 3 (*LAMB3*) in the lung (54)].

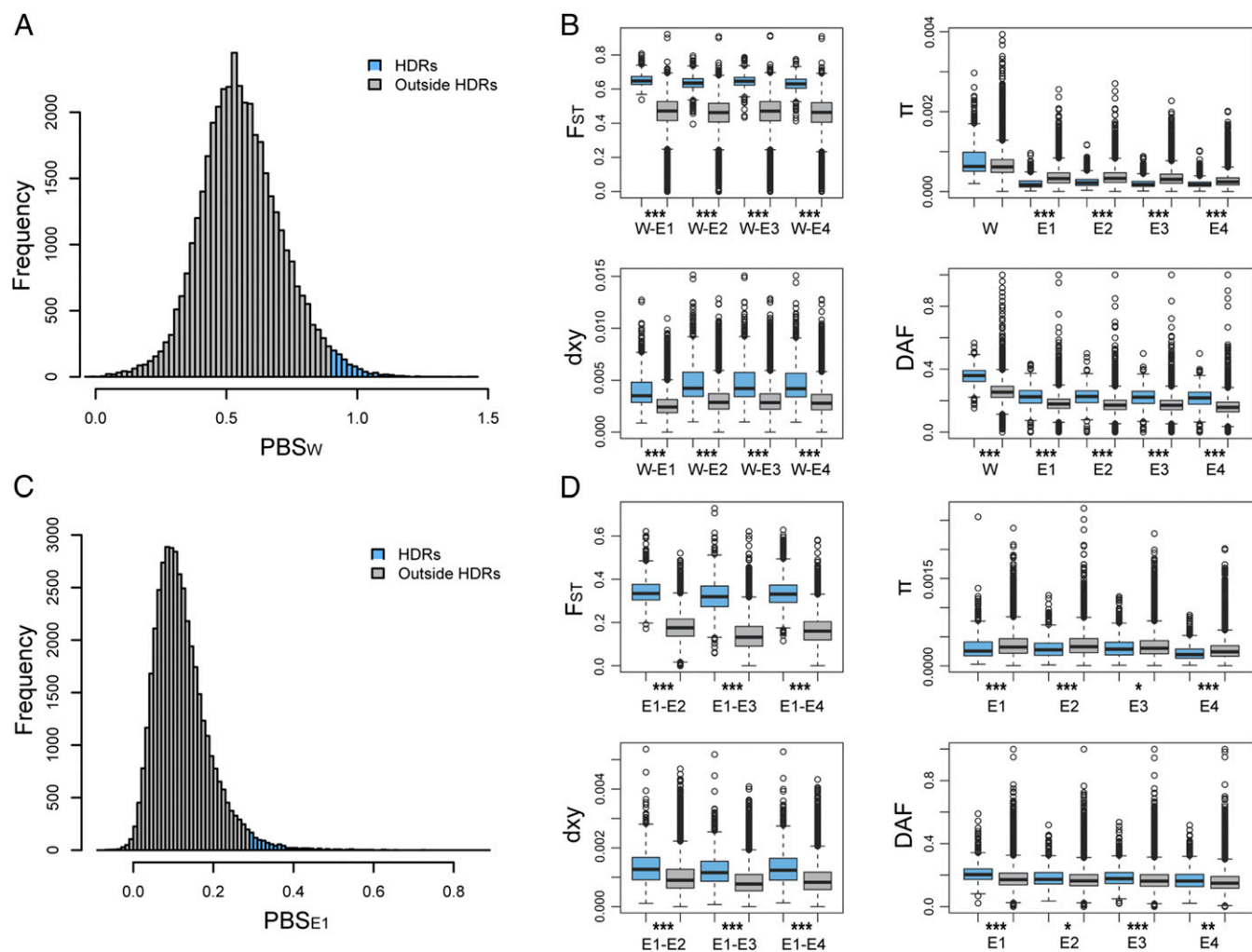
## Discussion

**Genomic Isolation and Speciation Between E and W Matrilines.** Population genomic analyses find substantial genomic isolation between W and E, with limited directional gene flow. Gene flow has been found only in one geographically restricted area, where no obvious geographic barrier (e.g., mountains) exists that could restrict dispersal. Further, the genetic patterns indicate that frogs from both E and W cross the YZR within their ranges and reject the null hypothesis of panmixia within a single species. Neutral processes, including divergence, changes of current, and ancestral effective population size ( $N_e$ ), and levels of gene flow have the potential to explain the broad pattern of differentiation between W and E. However, the HDRs between W and E exhibit significantly more genetic differentiation than our simulation using the inferred history. HDRs could result from positive selection and not demographic history (23). Furthermore, the HDRs in subpopulations of E exhibit significantly lower  $\pi$  and higher DAF compared with the rest of the genome (Fig. 5B), suggesting the action of positive selection on E. Because the HDRs of W and E harbor many genes that relate to GO categories involved in fertilization, particularly to spermatogenesis, they likely lead to reproductive isolation by suppressing gene flow. Studies of other species have detected the rapid evolution

of male reproductive genes, for instance, in primates (55), birds (19), fruit flies (56), and amphibians (57). Taken together with the fact that W, E2, and E4 occupy similar climatic environments (SI Appendix, Fig. S15) and display no differences in morphology, we suggest that endogenous selection is a dominant factor building reproductive isolation, resulting in speciation between W and E matrilineages.

Populations from the geographic region of admixture between E and W matrilineages occur along a narrow valley containing a tributary of the YZR. They have the mitogenome of E, whereas nuclear genes of W strongly dominate. Such mitonuclear discordance is common in vertebrates (58); however, here, a cohort of *N. parkeri* genomes offers an unusual opportunity to decipher the genomic pattern and the genetic mechanism of the phenomenon. Population genomic analyses reveal the E introgressed regions in hybrids are small and randomly dispersed in the genome. Those short introgressed segments suggest to us that they formed by hybridization, and were subsequently broken into even smaller sizes by recombination operating over many generations (59, 60). We detect no signal of selection in either the regions of introgression or the mitochondrial genome, and no significant enrichment of the pathways relevant to mitochondria or of the GOs of genes in introgressed regions. Thus, the mitonuclear discordance is not caused by natural selection driven by environmental factors (29, 61, 62). A likely scenario is establishment of reproductive isolating barriers (RIBs) resulting from secondary contact with historical and limited introgression after mid-Pleistocene divergence (63, 64). We hypothesize that hybrids formed historically only between females of E and males of W upon secondary contact, with hybrid females continuing to backcross to the males of W, leading to postzygotic, prezygotic, or both categories of RIBs.

**Ecological Adaptation Within E Matrilines.** E Tibetan frogs primarily segregate into high-elevation (>3,700 m, E2 and E4) and low-elevation (<3,700 m, E1 and E3) populations (Fig. 1A). This pattern likely reflects an ecological stratification in the southeastern QTP, including oxygen partial pressure (36), UV radiation (37), precipitation (38), and ambient temperature (39). If ecological selection is one of the forces that drives differentiation of the populations of E, one expects to find an imprint on genomic regions involved in ecological adaptation (27, 65).



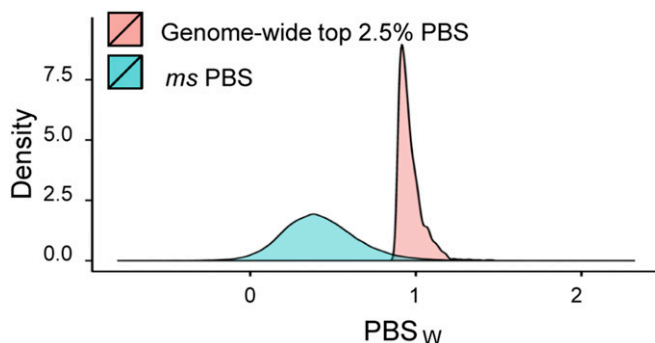
**Fig. 5.** Identification of HDRs. (A) Distribution of  $F_{ST}$ -based statistic of  $PBS_w$ . HDRs in top 2.5%  $PBS$  distribution, light blue; outside HDRs, gray. (B) Comparisons of HDRs of  $PBS_w$  in terms of  $F_{ST}$ ,  $\pi$ ,  $d_{xy}$ , and  $DAF$  versus the genomic background. (C) Distribution of  $PBS_{E1}$ . (D) Comparisons of HDRs of  $PBS_{E1}$  with the genomic background. Asterisks designate levels of significance between HDRs and outside HDRs by a two-tailed Mann–Whitney test (\* $P < 0.01$ ; \*\* $P < 1e-8$ ; \*\*\* $P < 2.2e-16$ ).

The HDRs of each subpopulation of E exhibit signals of selection compared with the background, including low  $\pi$ , increased  $d_{xy}$  between populations, and  $DAF$ . GO enrichment analyses point to environmental drivers of divergence. No clear geographic barriers separate the low-elevation sympatric pair E1 and E3. Although they were formed at different times (Fig. 3A), they share several GO functional categories, such as blood circulation and metabolic processes. Given the phylogeny and population history, these functions likely evolved independently in each lineage. The histological, physiological, and expressional evidence (Fig. 7) corresponds to the scenario that ecological stratification in the southeastern QTP has greatly promoted adaptive population diversification in situ.

In summary, *N. parkeri* is a useful model to investigate the processes and genetic bases of speciation along geographic and environmental gradients. We argue that natural selection plays important roles in driving continuing divergence within the species, and even in maintaining it. The extreme environments of the Tibetan Plateau can drive the rapid evolution of species [e.g., yak (33)]. Given the rapidity of changes and the challenging environment, the area is a natural laboratory for studying how selection drives adaptation; how environments influence evolutionary history; and, in some cases, how speciation can occur.

## Materials and Methods

**Sample Collection and Sequencing.** Tibetan frogs (45 in total) were selected for our genomic study based on matrilineal (mtDNA) genealogies (42). An additional 18 individuals come from the area of mixed clades in a tributary of the YZR (localities 25, 26, and 27; Fig. 1). One individual of *Nanorana pleskei* was used as the outgroup taxon (66). All collections were made according to



**Fig. 6.** Distribution of observed genome-wide top 2.5%  $PBS_w$  values compared with the simulated  $PBS_w$  values under the full model.



Category	Term	No. of genes	P value
Cluster 1	GO:0048538~thymus development	5	0.00
	GO:0048534~hemopoietic or lymphoid organ development	12	0.01
Cluster 2	GO:0002520~immune system development	12	0.01
	GO:0003006~reproductive developmental process	15	0.00
	GO:0048610~reproductive cellular process	11	0.00
	GO:0048609~reproductive process in a multicellular organism	18	0.01
	GO:0032504~multicellular organism reproduction	18	0.01
	GO:0019953~sexual reproduction	17	0.01
	GO:0007281~germ cell development	7	0.01
	GO:0007276~gamete generation	15	0.01
	GO:0048232~male gamete generation	12	0.02
	GO:0007283~spermatogenesis	12	0.02
Cluster 3	GO:0035270~endocrine system development	6	0.01
	GO:0030325~adrenal gland development	3	0.01

Annotation clusters with an enrichment score of  $\geq 2$  are shown.

animal use protocols approved by the Kunming Institute of Zoology Animal Care and Ethics Committee.

Total genomic DNA was extracted from liver, muscle, toe clips, or tadpole samples using the phenol/chloroform method (67). For each individual, 1–3 µg of DNA was sheared into fragments of 300–800 bp using the Covaris system. DNA fragments were processed and sequenced using Illumina paired-end sequencing technology (Illumina, Inc.). The 45 nonhybrid individuals and the outgroup taxa were sequenced to a target depth of 15×, and the 18 hybrid individuals were sequenced to target depth of 5×. The raw sequence data from this study have been submitted to the Genome Sequence Archive ([gsa.big.ac.cn/](https://www.genome.gov.cn/)) under accession nos. CRA000919 (*N. parkeri*) and CRA000918 (*N. pleskei*).

**SNP Calling and Filtering.** Raw sequence reads of each individual were mapped to the Tibetan frog reference genome (45) using BWA-ALN (v.0.7.4) (68) with default parameters. SAMtools (v.0.1.18) was used for sorting and removing PCR duplicates (69). To minimize false-positive SNP calls around indels, local realignment around indels was performed using the Genome Analysis Tool Kit (v.2.6-5) (70). Raw SNPs were extracted using SAMtools on the locally realigned BAM files with the command “samtools mpileup -a 20 -O 20 -C 50 -uDEF”.

To obtain high-quality genotype calls for downstream analyses, we kept SNPs that met the following criteria: (i) sites were at least 5 bp away from a predicted insertion/deletion, (ii) the consensus quality was  $\geq 40$ , (iii) sites did not have triallelic alleles and indels, (iv) the depth ranged from 2.5 to 97.5% in depth quartile, (v) SNPs had minor allele frequencies  $\geq 0.01$ , and (vi) SNPs occurred in more than 95% of high-coverage individuals (nonhybrids) and 75% of medium-coverage individuals (hybrids). The filtered data were phased using Beagle v.3.3.2 (71).

**Population Structure, Phylogenetic Inference, and Admixture Analyses.** We used PCA and population structure analysis to evaluate the genetic structuring of frogs. SNPs in scaffolds longer than 500 kb were extracted; they occupied about 76% of the entire genome. PCA was performed using the package GCTA (v.1.24.2) (72). The genomic ancestry of each individual was inferred using Frappe (v.1.1) (73). To avoid the effect of linkage disequilibrium, we selected one SNP for each interval of 50 kb. The postulated number of ancestral clusters (K) was set from two to five, and the maximum number of expectation-maximization iterations was set to 10,000.

Phylogenetic relationships were inferred via both coalescent and concatenation methods. To minimize the effects of potential alignment errors and regions with strong natural selection, we performed these analyses on putatively neutral genomic regions by filtering out the positions with repeat sequences, exons, and the 10 kb flanking them on each side. We randomly selected 1,000 neutral loci with a window size of 100 kb from the intergenic region. First, we reconstructed individual gene trees for each window based on the maximum likelihood (ML) approach using RAxML (v.8.1.15) (74). Support values for each node were inferred using 100 rapid bootstrap replicates based on the GTRGAMMA model. Second, for gene tree-based coalescent analysis, species trees were generated using MP-EST (75) and STAR (76, 77) and using population as the units of the tree tips. Third, for the concatenation analysis, ML trees were constructed based on the GTRGAMMA model using the concatenated sequences from the same set.

To infer admixture events, we applied the D-statistic using the qpDstat module in the ADMIXtools package (78). We also adopted a tree-based approach, which was implemented in TreeMix (79), to verify the existence of gene flow through modeling the migration values set from 0 to 4 with a block of 5,000 SNPs.

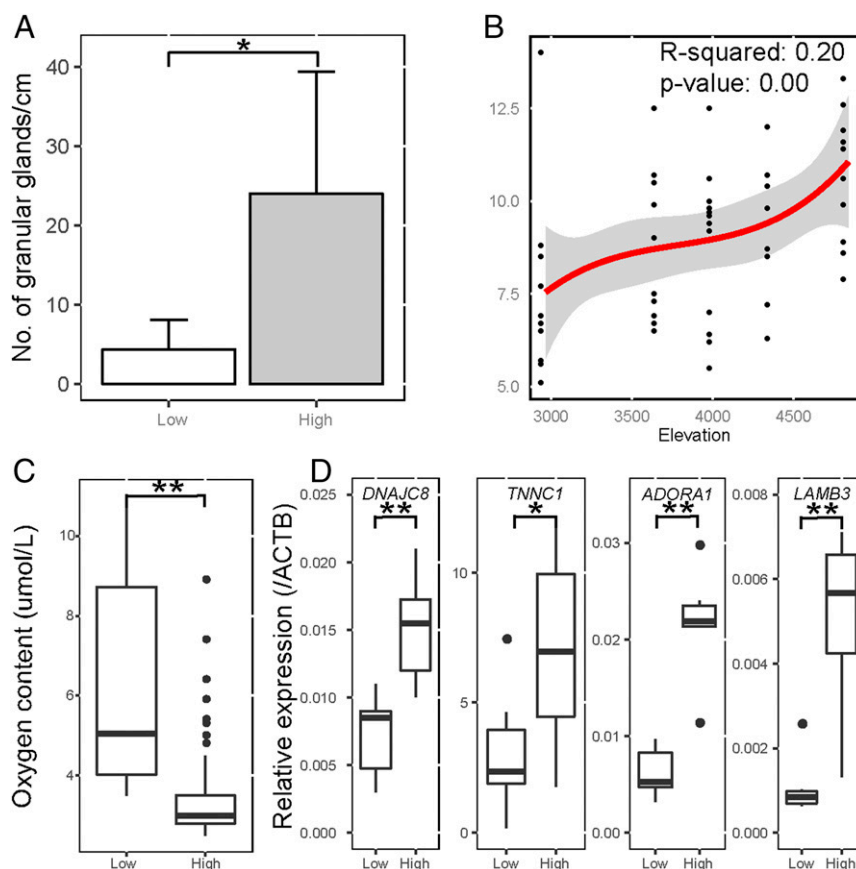
**Demographic Analysis Using G-PhoCS and PSMC.** G-PhoCS (46) was employed to infer the complete demographic history for *N. parkeri*, including population divergence times, ancestral population size, and migration rates based on 1,000 neutral loci (80). The parameters were inferred in a Bayesian manner using Markov Chain Monte Carlo to jointly sample model parameters and genealogies of the input loci (46). Migration scenarios were added by combining results of D-statistic tests, Frappe, and TreeMix because G-PhoCS often have limited ability to characterize complex migration scenarios (81). Additionally, two postdivergence migration bands were added to test if gene flow occurred between W and E since they separated. Each Markov chain was run for 2,000,000 generations while sampling parameter values every 20th iteration. Burn-in and convergence of each run were determined with TRACER 1.5 (82). More information about the control file of G-PhoCS is provided in [SI Appendix, section 1.2](#). Divergence times in units of years, effective population sizes, and migration rates were calibrated by the estimates of generation time and neutral mutation rate from previous studies (45, 83). We repeated the G-PhoCS analysis with four separate runs to obtain reliable and stable estimates for the demographic parameters. To validate the G-PhoCS inferences and test if the differential gene flow to W was correctly identified, we used *ms* (84) to perform simulations under different migration scenarios between W and E ([SI Appendix, section 1.3](#)). We used the inferred demographic parameters to produce 50 Mbp of sequences and compared the observed patterns of differentiation for W-E1 with those under different migration scenarios.

The trajectory of demographic histories for the five populations of Tibetan frogs was inferred by the PSMC model (85). Because PSMC has high false-negative rates at low sequence coverage, we restricted this analysis to the individual in each group with highest coverage ( $\geq 15\times$ ). In addition, a correction factor ( $-N$ ) was invoked to correct for the false-negative rate caused by the shallow sequence depth (80). The PSMC analysis was set as the following parameters: `-N25 -t15 -r5 -b -p "4+25 * 2+4+6"`. A bootstrapping approach with 100 replicates was performed to assess the variation in the inferred  $N_e$  trajectories. A generation time of 5 y and a neutral mutation rate of 0.776e-09 per site per year were used to convert the population sizes and scaled time into real sizes and time (45, 83).

The program fastsimcoal2 (86) was used to estimate the extent of population expansion within 30,000 y based on the model of G-PhoCS. Fifty time simulations were performed, and the result with the highest likelihood was kept. For each run, demographic estimates were obtained from 100,000 simulations (-n 100,000) and 40 expectation/conditional maximization cycles (-L 40) per parameter file.

**Inference of Mitochondrial Phylogeny and Hybrid Ancestry Assignment.** We used BWA-ALN (v.0.7.4) to map all raw sequence reads from each individual to the previously assembled complete mitochondrial genomes of *N. pleskei* (87), the closest relative of our Tibetan frog. SNPs were identified and filtered based on the same nuclear genome. The matrilineal (mitochondrial) genealogy was inferred using the ML method based on the GTRGAMMA model implemented in





**Fig. 7.** Morphological and physiological changes associated with elevation. (A) Bar plots of the differences in numbers of granular glands in the middorsal skin between low-elevation (2,968 m, E1) and high-elevation (4,859 m, E4) populations (two-tailed test:  $P < 0.05$ ). (B) Hb levels (grams per deciliter) at different elevations. Red lines show the best-fit regression line based on a third-order polynomial equation. The 95% confidence interval is shown in gray. (C) TOC ( $\mu\text{mol/L}$ ) in low (E1) and high (E4) elevations. (D) Expression level of *DNAJC8* in the brain, *TNNC1* and *ADORA1* in the heart, and *LAMB3* in the lung from low-elevation (E1) and high-elevation (E4) populations of E, respectively. Eight replicates were performed for each group. Statistically significant differences in differential expression are indicated by asterisk(s) (two-tailed  $t$  test:  $*P < 0.05$ ;  $**P < 0.01$ ).

RaxML (v.8.1.15) (74). PCAdmix (88) was used to identify the ancestries in each of the 18 hybrid individuals. We used W and E4 as the ancestral populations because they were the geographically closest populations and showed the highest signal of admixture in the f3 test. To prevent high-linkage blocks from having excessive influence on the inferred ancestry of a region, SNPs were thinned with  $r^2 > 0.80$  in all ancestral and admixed groups. PCA was performed using a window of 40 SNPs. A default calling threshold of 0.9 was used as the criterion to assign ancestry. However, systematic errors, such as incomplete lineage sorting and estimation errors, may have contributed to similar introgression signals in the PCAdmix analysis. To estimate these, we reanalyzed the PCAdmix analysis using non-E4 individuals as hybrids.

**Inference HDRs.** Genomic differentiation was calculated based on the PBS under the topological transformation described previously (89). The PBS value estimates the amount of sequence change along a population branch since its divergence from other branches of a population tree. First, pairwise  $F_{ST}$  statistics were calculated using Weir and Cockerham's method (90), implemented in VCFtools v0.1.11 (91), with nonoverlapping 50-kb genomic windows. Negative values of  $F_{ST}$  were treated as 0. Windows less 20 kb in length were excluded for further analysis. A log-transformation  $F_{ST}$  was used for the PBS calculation (92). The length of the branch leading to W since the divergence from all subpopulations of E was estimated as follows:

The length of the branch leading to E1 since the divergence from the remaining subpopulations of E clades was then estimated as follows:

$$PBS_{E1} = \frac{2T_{(E1,E2)} + T_{(E1,E3)} + T_{(E1,E4)} - T_{(E2,E3)} - T_{(E2,E4)}}{4}.$$

Calculations for branch lengths leading to subpopulations E2, E3, and E4 were similar to the equation for E1. HDRs were defined as the upper 2.5% of each PBS distribution. To avoid the potential assembly or mapping errors at the ends of scaffolds, we further removed high divergence peaks less than two consecutive windows in both ends of scaffolds in both sides.

**Characterization of HDRs in Terms of dxy, df,  $\pi$ , and DAF.** In addition to F<sub>ST</sub> and PBS, we applied several population genomic parameters to quantify and compare the outlier windows with the background genome, including the following: dxy between populations (93), df between populations (19), within-population  $\pi$  level (19, 23, 91), DAF, and population-scaled recombination rates ( $\rho$ ). We treated *N. pleskei* as the outgroup when comparing differences in DAF between W and E, and treated three nonadmixture W individuals (Fig. 2) as the outgroup when comparing DAF differences among subpopulations of E.

$$PBS_{West} = \frac{3T_{(West, E1)} + T_{(West, E2)} + T_{(West, E3)} + T_{(West, E4)} - T_{(E1, E2)} - T_{(E1, E3)} - T_{(E1, E4)}}{6}.$$

**GO Enrichment Analysis.** GO enrichment analysis was performed using DAVID (Database for Annotation Visualization and Integrated Discovery) (94). GO terms with less than two genes were excluded from further analysis. GO terms with a *P* value <0.05 were considered to be significantly enriched.

**Phenotype Data from Physiological and Morphological Detection.** Hb levels and tissue oxygen content (TOC) were collected in the field during June and July 2015. Measurements were taken in five communities in E along an elevation gradient: Nyingchi (29.61°N, 94.36°E, 2,968 m above sea level), Lhasa (29.67°N, 90.88°E, 3,671 m above sea level), Zhaxigang (29.75°N, 91.95°E, 4,011 m above sea level), Riduo (29.70°N, 92.23°E, 4,368 m above sea level), and Milashankou (29.80°N, 92.34°E, 4,859 m above sea level). At least 10 individuals were measured for each community. Hb level in peripheral blood was measured from 53 adult frogs, and muscular oxygen content was measured from 69 adult frogs. Hb was determined using Mission Plus Hb, immediately after drawing blood from the heart ventricle. TOC was determined via a fiber optic cable (PreSens Precision Sensing GmbH). We also made histological sections of the middorsal skin of seven frogs from the communities of Nyingchi and Milashankou, which represented populations from E at low and high elevations, respectively. The tissues were fixed in Heidenhain's Susa fixative to make paraffin sections (5 μm thick).

**RNA Isolation and Reverse Transcriptase PCR Assay.** Total RNAs from populations at low (Nyingchi, 2,968 m) and high (Milashankou, 4,859 m) elevations of E were extracted using the TRIzol total RNA extract kit (Tiangen). Reverse transcription was carried out using the Fermentas RevertAid First-Strand cDNA synthesis kit (Fermentas) to prepare templates for real-time quantitative PCR. The primers of candidate genes used for real-time PCR are displayed in *SI Appendix, Table S14*. Actin, Beta (*ACTB*) was used as a loading control.

**ACKNOWLEDGMENTS.** We thank Michael W. Nachman (University of California, Berkeley), Weiwei Zhai (Genome Institute of Singapore), and two reviewers for helpful comments and suggestions. We also thank Na Lin of Anhui University for assistance in file preparation. This work was supported by the Strategic Priority Research Program (B) (Grant XDB13020200 to J.C.) of the Chinese Academy of Sciences (CAS), National Natural Science Foundation of China (Grants 91431105 and 31622052 to J.C.), and the Animal Branch of the Germplasm Bank of Wild Species of CAS (Large Research Infrastructure Funding) (J.C.). W.-W.Z. was supported by Strategic Priority Research Program (B) (Grant XDB03030107) of CAS and the National Key Research and Development Program of China (Grant 2017YFC0505202). G.-D.W. was supported by the 13th Five-year Informatization Plan of CAS (Grant No. XXH13503-05). J.C. and G.-D.W. are supported by the Youth Innovation Promotion Association, CAS.

- Seehausen O, et al. (2014) Genomics and the origin of species. *Nat Rev Genet* 15: 176–192.
- Coyne J, Orr H (2004) *Speciation* (Sinauer, Sunderland, MA).
- Mayr E (1999) *Systematics and the Origin of Species from the Viewpoint of a Zoologist* (Harvard Univ Press, Cambridge, MA).
- Dobzhansky T (1937) *Genetics and the Origin of Species* (Columbia Univ Press, New York).
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci* 359:183–195, discussion 195.
- Avise JC (2001) *Phylogeography: The History and Formation of Species* (Harvard Univ Press, Cambridge, MA).
- Funk DJ (1998) Isolating a role for natural selection in speciation: Host adaptation and sexual isolation in *Neochlamis bebbianae* leaf beetles. *Evolution* 52:1744–1759.
- Schluter D (2000) *The Ecology of Adaptive Radiation* (Oxford Univ Press, Oxford).
- Nosil P (2012) *Ecological Speciation* (Oxford Univ Press, Oxford).
- Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends Ecol Evol* 16:330–343.
- Nosil P, Feder JL (2012) Genomic divergence during speciation: Causes and consequences. *Philos Trans R Soc Lond B Biol Sci* 367:332–342.
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends Genet* 28:342–350.
- Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P (2013) Geographic mode of speciation and genomic divergence. *Annu Rev Ecol Syst* 44:73–97.
- Nosil P (2008) Speciation with gene flow could be common. *Mol Ecol* 17:2103–2106.
- Pinho C, Hey J (2010) Divergence with gene flow: Models and data. *Annu Rev Ecol Syst* 41:215–230.
- Mayr E (1966) *Animal Species and Evolution* (Belknap Press of Harvard Univ Press, Cambridge, MA).
- Nolte AV, Tautz D (2010) Understanding the onset of hybrid speciation. *Trends Genet* 26:54–58.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18:375–402.
- Ellegren H, et al. (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491:756–760.
- Poelstra JW, et al. (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410–1414.
- Burri R, et al. (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res* 25:1656–1665.
- Lamichanay S, et al. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375.
- Malinsky M, et al. (2015) Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 350:1493–1498.
- Jones FC, et al. Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
- Phifer-Rixey M, Bomhoff M, Nachman MW (2014) Genome-wide patterns of differentiation among house mouse subspecies. *Genetics* 198:283–297.
- Carneiro M, et al. (2014) The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genet* 10:e1003519.
- Campbell-Staton SC, et al. (2017) Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. *Science* 357:495–498.
- Bragg JG, et al. (2018) Phylogenomics of a rapid radiation: The Australian rainbow skinks. *BMC Evol Biol* 18:15.
- Cheviron ZA, Brumfield RT (2009) Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution* 63:1593–1605.
- Cheviron ZA, Connaty AD, McClelland GB, Storz JF (2014) Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: Transcriptomic plasticity and thermogenic performance. *Evolution* 68:48–62.
- Storz JF, Scott GR, Cheviron ZA (2010) Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *J Exp Biol* 213:4125–4136.
- Favre A, et al. (2015) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol Rev Camb Philos Soc* 90:236–253.
- Qiu Q, et al. (2012) The yak genome and adaptation to life at high altitude. *Nat Genet* 44:946–949.
- Qu Y, et al. (2013) Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun* 4:2071.
- Simonson TS, et al. (2010) Genetic evidence for high-altitude adaptation in Tibet. *Science* 329:72–75.
- Wang D (2003) The geography of aquatic vascular plants of Qinghai Xizang (Tibet) Plateau. PhD dissertation (Wuhan University, Wuhan, China).
- Norsang G, Kocbach L, Stammes J, Tsoja W, Pincuo N (2011) Spatial distribution and temporal variation of solar UV radiation over the Tibetan Plateau. *Appl Phys Res* 3:37.
- Li SC, et al. (2007) Change of annual precipitation over Qinghai-Xizang Plateau and sub-regions in recent 34 Years. *J Desert Res* 27:307–314.
- Ma X, Lu X (2010) Annual cycle of reproductive organs in a Tibetan frog, *Nanorana parkeri*. *Anim Biol Leiden Neth* 60:259–271.
- Zhou W, et al. (2013) River islands, refugia and genetic structuring in the endemic brown frog *Rana kukunoris* (Anura, Ranidae) of the Qinghai-Tibetan Plateau. *Mol Ecol* 22:130–142.
- Kunming Institute of Zoology (CAS) (2018) AmphibiaChina. The database of Chinese amphibians. Available at [www.amphibiachina.org/](http://www.amphibiachina.org/). Accessed May 5, 2018.
- Zhou WW, et al. (2014) DNA barcodes and species distribution models evaluate threats of global climate changes to genetic diversity: A case study from *Nanorana parkeri* (Anura: Dicroglossidae). *PLoS One* 9:e103899.
- Liu J, et al. (2015) Phylogeography of *Nanorana parkeri* (Anura: Ranidae) and multiple refugia on the Tibetan Plateau revealed by mitochondrial and nuclear DNA. *Sci Rep* 5:9857.
- Fei L, Ye CY (2017) *Amphibians of China: I* (Science Press, Beijing).
- Sun YB, et al. (2015) Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc Natl Acad Sci USA* 112: E1257–E1262.
- Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43: 1031–1034.
- Zheng B, Xu Q, Shen Y (2002) The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: Review and speculation. *Quat Int* 35643:93–101.
- Rahmani Z, Maunoury C, Siddiqui A (1998) Isolation of a novel human voltage-dependent anion channel gene. *Eur J Hum Genet* 6:337–340.
- McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
- Xu X, Lai R (2015) The chemistry and biological activities of peptides from amphibian skin secretions. *Chem Rev* 115:1760–1846.
- Hershberger RE, et al. (2010) Coding sequence rare variants identified in MYBPC3, MYH6, TPM1, TNNI3 from 312 patients with familial or idiopathic dilated cardiomyopathy. *Circ Cardiovasc Genet* 3:155–161.
- Jenner TL, Rose-Meyer RB (2006) Loss of vascular adenosine A1 receptors with age in the rat heart. *Vascul Pharmacol* 45:341–349.
- Kaindl AM, et al. (2008) Erythropoietin protects the developing brain from hyperoxia-induced cell death and proteome changes. *Ann Neurol* 64:523–534.
- Kokkonen N, et al. (2007) Hypoxia upregulates carcinoembryonic antigen expression in cancer cells. *Int J Cancer* 121:2443–2450.
- Wyckoff GJ, Wang W, Wu CI (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature* 403:304–309.
- Ting CT, Tsaur SC, Wu ML, Wu CI (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282:1501–1504.

57. Malone JH, Michalak P (2008) Physiological sex predicts hybrid sterility regardless of genotype. *Science* 319:59.
58. Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* 21:3907–3930.
59. Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* 181:711–719.
60. Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet* 9:e1003521.
61. Irwin DE (2012) Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *Am Nat* 180:35–49.
62. Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. *Evolution* 62:1908–1920.
63. Chan KM, Levin SA (2005) Leaky prezygotic isolation and porous genomes: Rapid introgression of maternally inherited DNA. *Evolution* 59:720–729.
64. Colliard C, et al. (2010) Strong reproductive barriers in a narrow hybrid zone of West-Mediterranean green toads (*Bufo viridis* subgroup) with Plio-Pleistocene divergence. *BMC Evol Biol* 10:232.
65. Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14:671–688.
66. Che J, et al. (2010) Spiny frogs (Paini) illuminate the history of the Himalayan region and Southeast Asia. *Proc Natl Acad Sci USA* 107:13765–13770.
67. Sambrook J, Russell DW (2012) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Plainview, NY).
68. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
69. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
70. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.
71. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81:1084–1097.
72. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 88:76–82.
73. Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: Analytical and study design considerations. *Genet Epidemiol* 28:289–301.
74. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
75. Liu L, Yu L, Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol* 10:302.
76. Liu L, Yu L, Pearl DK, Edwards SV (2009) Estimating species phylogenies using coalescence times among sequences. *Syst Biol* 58:468–477.
77. Shaw TI, Ruan Z, Glenn TC, Liu L (2013) STRAW: Species TRee analysis web server. *Nucleic Acids Res* 41:W238–W241.
78. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093.
79. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967.
80. Wang GD, et al. (2016) Out of southern East Asia: The natural history of domestic dogs across the world. *Cell Res* 26:21–33.
81. Freedman AH, et al. (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet* 10:e1004016.
82. Rambaut A, Drummond AJ (2007) Tracer 1.5. Available at [tree.bio.ed.ac.uk/software/tracer/](http://tree.bio.ed.ac.uk/software/tracer/). Accessed July 3, 2011.
83. Ma X, Lu X (2009) Sexual size dimorphism in relation to age and growth based on skeletochronological analysis in a Tibetan frog. *Amphibia-Reptilia* 30:351–359.
84. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
85. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
86. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9:e1003905.
87. Chen G, Wang B, Liu J, Xie F, Jiang J (2011) Complete mitochondrial genome of *Nanorana pleskei* (Amphibia: Anura: Dicroglossidae) and evolutionary characteristics. *Curr Zool* 57:785–805.
88. Brisbin A, et al. (2012) PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* 84:343–364.
89. Lachance J, Tishkoff SA (2014) Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *Am J Hum Genet* 95:408–420.
90. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
91. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
92. Yi X, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78.
93. Ai H, et al. (2015) Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet* 47:217–225.
94. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.